Council of the European Union

Brussels, 17 October 2023
(OR. en)

13921/23

Interinstitutional File:
2021/0106(COD)

LIMITE

TELECOM 292
JAI 1276
COPEN 347
CYBER 235
DATAPROTECT 262
EJUSTICE 48
COSI 170
IXIM 183
ENFOPOL 415
RELEX 1154
MI 831
COMPET 962
CODEC 1798

NOTE

From:	Presidency
To:	Permanent Representatives Committee
No. prev. doc.:	13157/23
No. Cion doc.:	8115/21
Subject:	Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts - Preparation for the trilogue

INTRODUCTION

1. The Commission adopted the proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) on 21 April 2021.

2. The Council unanimously adopted its General Approach on the proposal on 6 December 2022.

3. The European Parliament (hereinafter: the EP) confirmed its position in a plenary vote on 14 June 2023.

4. Also on 14 June 2023, immediately after the vote in the EP, the co-legislators and the Commission held the first political trilogue on the AI Act, during which all three institutions outlined their priorities for the negotiations and the technical level was given a broad mandate to work on the entire proposal.

5. On 18 July 2023 the second political trilogue was held, during which some of the less controversial parts of the proposal were agreed and compromise was found on most elements of the chapter concerning measures in support of innovation.

6. On 2 and 3 October 2023 the third political trilogue took place in Strasbourg, during which further less controversial parts of the proposal were endorsed. In addition to this, the intention of the Presidency was to come to an agreement with the EP on three more controversial topics, namely the mechanism for the classification of AI systems as high-risk, the list of high-risk AI use cases, as well as subject matter and scope. However, talks on these topics proved inconclusive and no agreements were reached.

7. Since then four technical meetings with the EP have taken place, in order to prepare the next batch of provisions for agreement or for close alignment at the political level during the fourth political trilogue, which will take place on 24 October 2023 in Brussels.

8. Between 4 and 12 October 2023, the Presidency consulted the delegations, both during the meetings of the Working Party on Telecommunications and Information Society (hereinafter: WP TELECOM), and informally, on the compromise proposals and potential landing zones for agreement that are going to be discussed during the fourth political trilogue on 24 October 2023, and it has addressed all concerns expressed by the delegations, with regard to both the substance and the process for further negotiations.

II. POLITICAL ISSUES FOR POTENTIAL AGREEMENT DURING THE FOURTH TRILOGUE

9. The co-legislators intend to discuss the following three topics with a view to reaching a provisional agreement at political level during the fourth trilogue:

- **Classification of AI systems as high-risk** (Recitals 32 and 32a, Articles 6 and 7, Article 51, Article 65a *(new)*, subparagraph 2 of Annex III *(new)* and Section D of Annex VIII *(new)*)

- **List of high-risk AI use cases** (Annex III) *[with the exception of use cases related to biometrics and law enforcement authorities, which will be discussed together with Article 5 on prohibitions at a later stage]*

- **Testing high-risk AI systems in real world conditions outside AI regulatory sandboxes** (Articles 54a and 54b)

10. As regards the first two topics above (**classification of AI systems as high-risk** and **the list of high-risk AI use cases**), the Presidency obtained a revised mandate and some flexibilities from the Permanent Representatives Committee on 29 September, and the intention is to approach the negotiations with the EP on 24 October 2023 based on that mandate (document 13157/23, Section II of the Annex, points A and B).

11. Concerning the third topic listed above, namely **testing high-risk AI systems in real world conditions outside AI regulatory sandboxes** (Articles 54a and 54b), the compromise proposal, as set out in Annex I to this note, is based on elements taken from the General Approach of the Council from 6 December 2022, but it contains some additional safeguards which the Presidency considers necessary in order to reach an agreement with the EP on this topic. More specifically, the modifications include the following elements:

- the requirement for approval from market surveillance authority to conduct testing in real world conditions has been added, with tacit approval possible after 45 days;
- in the case of testing in real world conditions in the areas of law enforcement, migration, asylum and border control management, a proposal for registration in a non-publicly available database has been included;
- the duration of testing in real world conditions has been set as six months, but it could be extended for a further six more months, subject to prior notification by the provider to the market surveillance authority;
- as there are cases where law enforcement authorities might not be in a position to obtain informed consent from affected persons before testing in real world conditions, a

requirement has been added for these authorities to include in their real world testing plan an analysis on why these persons wouldn't be negatively impacted;

- the right for affected persons to request to delete their data after testing in real world conditions has been explicitly added;
- the right for market surveillance authorities to request information related to testing in real world conditions from providers and prospective providers has been included, including the power to conduct inspections if needed.

Delegations are asked to indicate whether they would be open to the possibility of adding the additional safeguards to real world testing of high-risk AI systems, as described above, and as set out in Annex I to this note.

III. TOPICS FOR FOR HIGH-LEVEL POLITICAL GUIDANCE AIMED AT ALIGNING THE POSISTIONS OF CO-LEGISLATORS DURING THE FOURTH TRILOGUE

12. Apart from reaching a provisional agreement on the issues listed in point II above, the co-legislators intend to discuss the following two topics with a view to aligning their positions during the fourth trilogue:

- **Foundation models/General purpose AI systems and governance**

- **Prohibitions, law enforcement and national security**

In Annex II to this note, delegations will find a description of potential landing zones with regard to these two topics, prepared at technical level, but without concrete drafting proposals. The intention of the Presidency is to discuss these topics with the EP during the upcoming fourth trilogue, based on these landing zone proposals, in order to reach a common understanding on how the final solutions could be drafted. With this understanding, the technical level would be tasked to prepare concrete drafting compromise proposals, which would be then submitted to the Permanent Representatives Committee with a request for a revised mandate ahead of the subsequent trilogue planned for 6 December 2023.

In this context, delegations are requested to provide their early feedback and views on the landing zones as presented in sections I and II of Annex II to this note.

The Presidency will use this feedback as the basis for the discussions during the fourth trilogue, and for the drafting of final compromise proposals at technical level.

V. CONCLUSION

13. In light of the above, and with a view to obtaining a revised mandate for trilogue negotiations on the AI Act on 24 October 2023, the Permanent Representatives Committee is invited to:

– **indicate their flexibility with regard to the question presented in Part II of this note,**

– **provide the Presidency with early feedback on the landing zones concerning the topics referred to in Part III of this note.**

Articles 54a and 54b - Testing of high-risk AI systems in real world conditions outside AI regulatory sandboxes

	Commission Proposal	EP Mandate	Council Mandate		Presidency2	
Article 54(2a)						
R	541b				***Article 54a Testing of high-risk AI systems in real world conditions outside AI regulatory sandboxes***	R
Article 54(2b), first subparagraph						
R	541c				***1. Testing of AI systems in real world conditions outside AI regulatory sandboxes may be conducted by providers or prospective providers of high-risk AI systems listed in Annex III, in accordance with the provisions of this Article and the real-world testing plan referred to in this Article.***	R
Article 54(2b), second subparagraph						
R	541d				***The detailed elements of the real-world testing plan shall be specified in implementing acts adopted by the Commission in accordance with the examination procedure referred to in Article 74(2).***	R
Article 54(2c)						

	Commission Proposal	EP Mandate	Council Mandate	Presidency2	
R	541e			***This provision shall be without prejudice to Union or Member State legislation for the testing in real world conditions of high-risk AI systems related to products covered by legislation listed in Annex II.***	***2c. This provision shall be without prejudice to Union or Member State legislation for the testing in real world conditions of high-risk AI systems related to products covered by legislation listed in Annex II.*** R
	Article 54(2d)				
R	541f			***2. Providers or prospective providers may conduct testing of high-risk AI systems referred to in Annex III in real world conditions at any time before the placing on the market or putting into service of the AI system on their own or in partnership with one or more prospective users.***	***2. Providers or prospective providers may conduct testing of high-risk AI systems referred to in Annex III in real world conditions at any time before the placing on the market or putting into service of the AI system on their own or in partnership with one or more prospective users.*** R
	Article 54(2e)				
R	541g			***3. The testing of high-risk AI systems in real world conditions under this Article shall be without prejudice to ethical review that may be required by national or Union law.***	***3. The testing of high-risk AI systems in real world conditions under this Article shall be without prejudice to ethical review that may be required by national or Union law.*** R
	Article 54(2f)				
R	541h			***4. Providers or prospective providers***	***4. Providers or prospective providers*** R

	Commission Proposal	EP Mandate	Council Mandate	Presidency2
			may conduct the testing in real world conditions only where all of the following conditions are met:	***may conduct the testing in real world conditions only where all of the following conditions are met:***
Article 54(2f), point (a)				
R	541i		***(a) the provider or prospective provider has drawn up a real-world testing plan and submitted it to the market surveillance authority in the Member State(s) where the testing in real world conditions is to be conducted;***	***(a) the provider or prospective provider has drawn up a real-world testing plan and submitted it to the market surveillance authority in the Member State(s) where the testing in real world conditions is to be conducted;***
Article 54(2f), point (b)				
R	541j		***(b) the market surveillance authority in the Member State(s) where the testing in real world conditions is to be conducted have not objected to the testing within 30 days after its submission;***	***(b) the market surveillance authority in the Member State(s) where the testing in real world conditions is to be conducted has approved the testing in real world conditions and the real-world testing plan. Where the market surveillance authority in that Member State has not provided with an answer in 45 days, the testing in real world conditions and the real-world testing plan shall be understood as approved;***
Article 54(2f), point (c)				

	Commission Proposal	EP Mandate	Council Mandate	Presidency2
R 541k			***(c) the provider or prospective provider with the exception of high-risk AI systems referred to in Annex III, points 1, 6 and 7 in the areas of law enforcement, migration, asylum and border control management, and high risk AI systems referred to in Annex III point 2, has registered the testing in real world conditions in the EU database referred to in Article 60(5a) with a Union-wide unique single identification number and the information specified in Annex VIIIa;***	***(c) the provider or prospective provider with the exception of high-risk AI systems referred to in Annex III, points 1, 6 and 7 in the areas of law enforcement, migration, asylum and border control management, and high risk AI systems referred to in Annex III point 2, has registered the testing in real world conditions in the EU database referred to in Article 60(5a) with a Union-wide unique single identification number and the information specified in Annex VIIIa. In the cases of high-risk AI systems referred to in Annex III, points 1, 6 and 7 in the areas of law enforcement, migration, asylum and border control management, the registration will take place in a non-publicly available area of the EU database, as referred to in Article 60 (xx);*** R
Article 54(2f), point (d)				
R 541l			***(d) the provider or prospective provider conducting the testing in real world conditions is established in the Union or it has appointed a legal***	***(d) the provider or prospective provider conducting the testing in real world conditions is established in the Union or it has appointed a legal*** R

	Commission Proposal	EP Mandate	Council Mandate	Presidency2	
			representative for the purpose of the testing in real world conditions who is established in the Union;	***representative who is established in the Union;***	
Article 54(2f), point (e)					
R	541m		***(e) data collected and processed for the purpose of the testing in real world conditions shall not be transferred to countries outside the Union, unless the transfer and the processing provides equivalent safeguards to those provided under Union law;***	***(e) data collected and processed for the purpose of the testing in real world conditions shall not be transferred to third countries outside the Union, unless the transfer and the processing provides equivalent safeguards to those provided under Union law;***	R
Article 54(2f), point (f)					
R	541n		***(f) the testing in real world conditions does not last longer than necessary to achieve its objectives and in any case not longer than 12 months;***	***(f) the testing in real world conditions does not last longer than necessary to achieve its objectives and in any case not longer than 6 months, which may be extended for an additional amount of 6 months, subject to prior notification by the provider to the market surveillance authority, accompanied by an explanation on the need for such time extension;***	R
Article 54(2f), point (g)					
R	541o		***(g) persons***	***(g) persons***	R

	Commission Proposal	EP Mandate	Council Mandate	Presidency2	
			belonging to vulnerable groups due to their age, physical or mental disability are appropriately protected;	***belonging to vulnerable groups due to their age, physical or mental disability are appropriately protected;***	
Article 54(2f), point (h)					
R	541p		***(h) where a provider or prospective provider organises the testing in real world conditions in cooperation with one or more prospective users, the latter have been informed of all aspects of the testing that are relevant to their decision to participate, and given the relevant instructions on how to use the AI system referred to in Article 13; the provider or prospective provider and the user(s) shall conclude an agreement specifying their roles and responsibilities with a view to ensuring compliance with the provisions for testing in real world conditions under this Regulation and other applicable Union and Member States legislation;***	***(h) where a provider or prospective provider organises the testing in real world conditions in cooperation with one or more prospective users, the latter have been informed of all aspects of the testing that are relevant to their decision to participate, and given the relevant instructions on how to use the AI system referred to in Article 13; the provider or prospective provider and the user(s) shall conclude an agreement specifying their roles and responsibilities with a view to ensuring compliance with the provisions for testing in real world conditions under this Regulation and other applicable Union and Member States legislation;***	R
Article 54(2f), point (i)					
R	541q		***(i) the subjects of the testing in real world conditions have given informed***	***(i) the subjects of the testing in real world conditions have given informed***	R

		Commission Proposal	EP Mandate	Council Mandate		Presidency2	
				consent in accordance with Article 54b, or in the case of law enforcement, where the seeking of informed consent would prevent the AI system from being tested, the testing itself and the outcome of the testing in the real world conditions shall not have a negative effect on the subject;		***consent in accordance with Article 54b, or in the case of law enforcement, where the seeking of informed consent would prevent the AI system from being tested, the testing itself and the outcome of the testing in the real world conditions shall not have any negative effect on the subject. In these cases, providers or prospective providers shall include in their real-world testing plan a detailed analysis on why the testing shall not have a negative effect on the subject.***	
	Article 54(2f), point (j)						
R	541r			***(j) the testing in real world conditions is effectively overseen by the provider or prospective provider and user(s) with persons who are suitably qualified in the relevant field and have the necessary capacity, training and authority to perform their tasks;***		***(j) the testing in real world conditions is effectively overseen by the provider or prospective provider and user(s) with persons who are suitably qualified in the relevant field and have the necessary capacity, training and authority to perform their tasks;***	R
	Article 54(2f), point (k)						
R	541s			***(k) the predictions, recommendations or decisions of the AI system can be effectively reversed***		***(k) the predictions, recommendations or decisions of the AI system can be effectively reversed***	R

	Commission Proposal	EP Mandate	Council Mandate	Presidency2
			or disregarded.	***and disregarded.***
Article 54(2g)				
R 541t			***5. Any subject of the testing in real world conditions, or his or her legally designated representative, as appropriate, may, without any resulting detriment and without having to provide any justification, withdraw from the testing at any time by revoking his or her informed consent. The withdrawal of the informed consent shall not affect the activities already carried out and the use of data obtained based on the informed consent before its withdrawal.***	***5. Any subject of the testing in real world conditions, or his or her legally designated representative, as appropriate, may, without any resulting detriment and without having to provide any justification, withdraw from the testing at any time by revoking his or her informed consent and request the immediate and permanent deletion of their personal data. The withdrawal of the informed consent shall not affect the activities already carried out.*** ***5a. Member States shall confer their market surveillance authorities the powers of requiring providers and prospective providers information, of carrying out unannounced remote or on-site inspections and on performing checks on the development of the testing in real world conditions and the related products. Market surveillance authorities shall use these powers to*** R

	Commission Proposal	EP Mandate	Council Mandate	Presidency2
				ensure a safe development of these tests.
Article 54(2h)				
541u			***6. Any serious incident identified in the course of the testing in real world conditions shall be reported to the national market surveillance authority in accordance with Article 62 of this Regulation. The provider or prospective provider shall adopt immediate mitigation measures or, failing that, suspend the testing in real world conditions until such mitigation takes place or otherwise terminate it. The provider or prospective provider shall establish a procedure for the prompt recall of the AI system upon such termination of the testing in real world conditions.***	***6. Any serious incident identified in the course of the testing in real world conditions shall be reported to the national market surveillance authority in accordance with Article 62 of this Regulation. The provider or prospective provider shall adopt immediate mitigation measures or, failing that, suspend the testing in real world conditions until such mitigation takes place or otherwise terminate it. The provider or prospective provider shall establish a procedure for the prompt recall of the AI system upon such termination of the testing in real world conditions.***
Article 54(2i)				
541v			***7. Providers or prospective providers shall notify the national market surveillance authority in the Member State(s) where the testing in real world conditions***	***7. Providers or prospective providers shall notify the national market surveillance authority in the Member State(s) where the testing in real world conditions***

	Commission Proposal	EP Mandate	Council Mandate	Presidency2	
			is to be conducted of the suspension or termination of the testing in real world conditions and the final outcomes.	***is to be conducted of the suspension or termination of the testing in real world conditions and the final outcomes.***	
Article 54(2j)					
R	541w		***8. The provider and prospective provider shall be liable under applicable Union and Member States liability legislation for any damage caused in the course of their participation in the testing in real world conditions.***	***8. The provider and prospective provider shall be liable under applicable Union and Member States liability legislation for any damage caused in the course of their participation in the testing in real world conditions.***	R
Article 54b					
R	541x		***Article 54b Informed consent to participate in testing in real world conditions outside AI regulatory sandboxes***	***Article 54b Informed consent to participate in testing in real world conditions outside AI regulatory sandboxes***	R
Article 54b(1), first subparagraph					
R	541y		***1. For the purpose of testing in real world conditions under Article 54a, informed consent shall be freely given by the subject of testing prior to his or her participation in such testing and after having been duly informed with concise, clear, relevant, and understandable information regarding:***	***1. For the purpose of testing in real world conditions under Article 54a, informed consent shall be freely given by the subject of testing prior to his or her participation in such testing and after having been duly informed with concise, clear, relevant, and understandable information regarding:***	R

	Commission Proposal	EP Mandate	Council Mandate	Presidency2	
Article 54b(1), second subparagraph					
R	541z		***(i) the nature and objectives of the testing in real world conditions and the possible inconvenience that may be linked to his or her participation;*** ***(ii) the conditions under which the testing in real world conditions is to be conducted, including the expected duration of the subject's participation;*** ***(iii) the subject's rights and guarantees regarding participation, in particular his or her right to refuse to participate in and the right to withdraw from testing in real world conditions at any time without any resulting detriment and without having to provide any justification;*** ***(iv) the modalities for requesting the reversal or the disregard of the predictions, recommendations or decisions of the AI system;*** ***(v) the Union-wide unique single identification number of the testing in real world conditions in accordance with***	***(i) the nature and objectives of the testing in real world conditions and the possible inconvenience that may be linked to his or her participation;*** ***(ii) the conditions under which the testing in real world conditions is to be conducted, including the expected duration of the subject's participation;*** ***(iii) the subject's rights and guarantees regarding participation, in particular his or her right to refuse to participate in and the right to withdraw from testing in real world conditions at any time without any resulting detriment and without having to provide any justification, as well as his or her right to ask for the permament deletion of his or her personal data used during the test;*** ***(iv) the modalities for requesting the reversal and the disregard of the predictions, recommendations or decisions of the AI system;*** ***(v) the Union-***	R

	Commission Proposal	EP Mandate	Council Mandate	Presidency2
			Article 54a(4c) and the contact details of the provider or its legal representative from whom further information can be obtained.	***wide unique single identification number of the testing in real world conditions in accordance with Article 54a(4c) and the contact details of the provider or its legal representative from whom further information can be obtained.***
Article 54b(2)				
R 541aa			***2. The informed consent shall be dated and documented and a copy shall be given to the subject or his or her legal representative.***	***2. The informed consent shall be dated and documented and a copy shall be given to the subject or his or her legal representative.*** R

ANNEX II

Section I – Foundation models/General purpose AI systems and governance

I. New rules for foundation models and general purpose AI (GPAI)

Although with different approaches, both Parliament and Council introduced rules to address concerns arising with the use of foundation models and GPAI.

The Council has made an important effort to address the problem at GPAI system level. This effort is most valuable for the sake of ensuring proper allocation of responsibility in the value chain when used at scale by downstream providers to develop high risk AI systems. This is why it seems very important to keep this proposal with in full respect of the risk based approach.

At the same time, the capabilities and complexity of foundation models are such that certain tailored transparency obligations are necessary to ensure that downstream providers can build AI systems (including general purpose AI systems) on foundation models in a way that is safe and compliant with the AI Act, minimising the risk to violate fundamental rights and safety. In addition, when foundation models have particularly high capabilities or emergent capabilities that are not yet fully understood, those models warrant heightened attention in view of possible systemic risk, such as risk to life or public health at large or negative effect on democratic processes through massive disinformation.

1. Introducing obligations for all foundation models

Foundation models differ from more traditional narrow AI models on the basis of their numerous capabilities. Learning objectives tend to be general and learning task independent. Therefore, a possible definition could refer to 'AI model that is capable to competently perform a wide range of distinctive tasks'. Concrete benchmarks for evaluating capabilities of these models in terms of tasks and competence to perform these tasks should be developed and set out in implementing acts.

All providers of foundation models should be subject to the following basic transparency obligations:

Before the foundation model is put on the market:

- documenting the model and training process, including the results of internal red teaming,
- carrying out and documenting model evaluation in accordance with standardised protocols and tools (i.e. benchmarks),

and after the foundation model is put on the market:

- providing information and documentation to the downstream provider, and
- enabling the testing of foundation models by downstream providers.

Providers of foundation models should collaborate with authorities (e.g. the Office[1]), who may, upon alert, request the disclosure of the documentation.

2. Introducing additional obligations for very capable foundation models

Providers of very capable foundation models (see definition below) should be subject to additional obligations, on top of the transparency obligations under 1:

Before the very capable foundation model is put on the market (or in case of retraining):

- regular external red-teaming through vetted red-testers (to be vetted by the Office), with a view to uncover vulnerabilities and identify areas for risk mitigation, the results of which need to be submitted to the Office,
- introducing a risk assessment and mitigation system, also covering possible systemic risks,

and after the very capable foundation models is on the market:

- regular compliance controls organised by the Office and carried out through independent auditors/researchers, which entails checking compliance with the obligations regarding transparency.

At this point in time, the knowledge and understanding of possible systemic risks of very capable foundation models and the tools to address them is still evolving. Therefore, to facilitate the implementation of the obligation to conduct risk assessment and mitigation, the Office should set up a forum of cooperation for providers of very capable foundation models to discuss relevant best practices and draw up a voluntary code of conduct. The code of conduct should be submitted to the Office and endorsed by the Commission. The Office should also be able to monitor the implementation of commitments under these codes of conduct.

Definition of very capable foundation models:

Very capable foundation models should be understood as foundation models whose capabilities go beyond the current state-of-the-art and may not yet be fully understood. Because there are not yet tools and methodologies to predict and measure the capabilities of those models, researchers have identified proxies such as the amount of compute used for the training (FLOPs). The amount of compute measures in FLOPs corresponding to expectations in relation to the most performant models expected to be released in 2024/25 could be the triggering threshold to capture the most advanced models of the future at the time when the AI Act enters into force. This proxy results from very recent research and calculations based solely on large language models and depends on several factors. As models become increasingly efficient, in order to ensure that the system remains futureproof, it would be essential that implementing acts can set out the concrete tools and methodologies to predict and measure the capabilities of those models. The FLOPs threshold would be updated or adapted when needed on the basis of these tools and methodologies, following stakeholder and expert consultation.

In light of the above, a workable mechanism could be to presume a foundation model is 'very capable' when the threshold of FLOPs is reached, triggering the obligation for the provider to notify

[1] The new rules would require a new approach to oversight and enforcement. It seems most appropriate to centralise the enforcement of the rules on foundation models at EU-level, notably through the Office (see section II).

the Office. However, a provider could rebut this presumption by demonstrating in the notification that the foundation model in question should not be considered very capable, despite reaching the threshold. At the same time, there should be the possibility for the Office to exceptionally consider a model as 'very capable' even below the threshold of FLOPs, notably following an investigation if this has been flagged by the scientific community.

Nevertheless, in order to have the most future proof criteria to define what is a very capable model, the following metrics could also be used: the amount of data consumed in training, which, according to some research, it is not expected to alter in the coming years, as well as on the potential impact of these foundational models in users, established by the amount of high-risk AI applications that are built on the basis of such foundation model. This would require that in the registration of high-risk AI systems, providers would specify whether the high risk AI system is built on a foundation model (and which one).

3. Introducing obligations for GPAI systems at scale

GPAI systems built on foundation models and used at scale in the EU may bring about higher risks due to their wide adoption. Therefore, providers of such GPAI systems should be subject to the following obligations:

- regular external red-teaming through vetted red-testers, with a view to uncover vulnerabilities and identify areas for risk mitigation, the results of which need to be submitted to the Office.
- introducing a risk assessment and mitigation system, also covering possible systemic risks.

Further consideration is needed on the interplay between risk mitigation at system level and at model level. While aiming to avoid disproportionate burden, especially on smaller actors, one could consider introducing certain risk management requirements for both model and system level, making use of the same regulatory tools, i.e. allowing providers to discuss best practices and encouraging the development of codes of conducts that set out how compliance with the risk management obligations can be achieved. Further consideration is also needed as to how to ensure inclusion of guardrails, at either (very capable) model or GPAI system (at scale) level, to address the risk of illegal or harmful output of the GPAI system, including safeguards against misuse or autonomous use to generate such output, especially where such outputs would generate systemic risks.

The determination when a GPAI systems is used at scale is based on the impact and reach. For this purpose, the relevant threshold should be an amount of [10 000] registered business users (i.e. developers) or [45 million] registered end users, as appropriate. The calculation is straight-forward for GPAI systems provided through an API. For GPAI systems provided through a 'library', a methodology needs to be developed. Implementing acts should set out the concrete methodology for the calculation of users. Providers of GPAI systems should have the possibility to request exemption from the obligations for GPAI systems used at scale despite meeting this threshold, if they can demonstrate that the GPAI system in question does not pose the specific risks associated with a wide adoption of GPAI systems. Conversely, it could be considered that obligations would be extended to providers of GPAI systems which do not reach the relevant scale thresholds, if they are shown to give rise to risks, including systemic risks, that cannot be adequately addressed at the level of the underlying (very capable) model.

4. Avoiding loopholes in the risk-based approach

Furthermore, to align with the risk-based approach, all providers of GPAI systems should explicitly state whether the GPAI can be used for high-risk uses or not. Providers who exclude use for high-risk purposes should introduce measures to detect and prevent such use. Providers who allow certain high-risk uses should make sure the GPAI system complies with the requirements for high-risk AI systems for each allowed high-risk use. The requirements may nevertheless be adapted via implementing acts in order to take into account the specificities of GPAI systems.

5. Introducing obligations to support enforcement of copyright protections

The EU Copyright Directive already foresees that right holders can opt-out from their content to be used for training foundation models (TDM exception). However, there is a need for targeted provisions that facilitate the enforcement of copyright rules in the context of foundation models. Providers of foundation models should demonstrate that they have taken adequate measures to ensure the models are trained in compliance with applicable Union copyright law, in particular respect the opt-out from the TDM exception. In addition, providers of foundation models should make publicly available a sufficiently detailed summary about the content used for training and information about their policies to manage copyright-related aspects. The Office should provide a template to ensure uniform application of this obligation.

6. Introducing obligations to ensure transparency of AI-generated content

Considering the possible impact on the information ecosystem, the obligations as regards transparency of AI-generated content should be reinforced. For this purpose, providers of AI systems that generate output, typically based on foundation models, should be obliged to ensure that the output is detectable as artificially generated or manipulated.

Providers should ensure their technical solutions are effective, interoperable, robust and reliable as far as this is technically feasible taking into account acknowledged state-of-the-art. Such provision should be formulated in a technology neutral way, as today there are not yet consolidated technical solutions for machine-generated text. The implementation should be specified through standards or voluntary codes of conducts, the latter maybe with a possibility for recognition by the Commission.

II. Governance of the new rules for foundation models and GPAI

The new rules for foundation models and GPAI require a new approach to the governance. While for AI systems the market surveillance system will apply, these new rules for foundation models and GPAI require a new system of oversight and enforcement. The complexity and capability of these models and systems are such that centralising expertise would be important.

1. Introducing centralised supervision of foundation models and GPAI at scale

The enforcement of the new rules on foundation models and GPAI at scale should take place on EU-level. Besides the enforcement, there is a need for a centralised structure for supervision,

monitoring and foresight, that develops an understanding of trends and potential risks of such models and systems, including risks of negative effects on health, safety and fundamental rights, as well as possible serious or systemic risks, such as risk to life or public health at large or negative effects on democratic processes.

For this purpose, the "AI Office" should be set-up as a new governance structure with the following specific tasks in respect of foundation models and GPAI:

- Enforcement and supervision of the new rules on foundation models and GPAI systems used at scale, including defining audit procedures and modalities as well as having powers to:
    - request documentation,
    - organise and carry out compliance controls for very capable foundation models and GPAI used at scale, for which it can involve independent auditors or experts, and
    - carry out investigations upon alerts and take corrective action, including suspension of the model as a last resort
- Monitoring potential serious risks of foundation models and GPAI, making recommendations and issuing warnings in case of identified risks
- Developing supportive tools, such as standardised protocols and tools for model evaluations and best practices for red teaming
- Setting up a forum of cooperation for providers of very capable foundation models and GPAI systems used at scale to discuss best practices for mitigation of serious risks
- Setting up a forum for collaboration with the open-source community with a view to foster cooperation and identify and develop best practices for the safe development and use of open-source foundation models
- Collaboration with scientific community, including setting up a registry of 'vetted red-testers' and a group of experts for scientific advice
- Collecting complaints from citizens and alerts about foundation models and GPAI
- Support in the context of international cooperation related to the enforcement of applicable rules to foundation models and GPAI and testing capabilities, such as the UK proposal for an international AI safety testing framework

There should be a strong link with the scientific community to support the enforcement. The Office would involve independent experts for audits and provide support to independent red-teaming. A key element would be a new scheme for 'vetted red-testers', building on the model of 'trusted-flaggers' under the DSA. For this purpose, the Office would provide vetting for independent experts and expert organisations with expertise and competence for carrying out red teaming. The Office should set up a register of these vetted red-testers, which providers of foundation models can refer to for external red-teaming.

Being the first body worldwide with powers to enforce rules on foundation models and GPAI, the Office would become an international reference point for AI governance. The Office could also support the EU and Member States in the context of international cooperation related to the enforcement of applicable rules to foundation models and GPAI.

2. Complementarity with the governance and enforcement mechanism for AI systems

The Office would be an additional building block that complements the governance and enforcement mechanism for AI systems. This means that the supervision of AI systems under the well-functioning market surveillance system is preserved, and the AI Board remains the coordination platform for national authorities and advisory body to the Commission.

However, the Office could possibly reinforce the existing governance with more general tasks:

- Support the Commission in examining complaints about the use of biometrics,
- Advising the Commission on matters related to the Regulation,
- Supporting national authorities in the implementation of the AI Act,
- Acting as a secretariat for the AI Board, and
- Contribute to the preparation of templates and tools for stakeholders.

3. Using synergies at EU-level

The Office should be a visible governance body that is administratively hosted within the Commission. It should be presented as a self-standing organisation, including through an online presence that allows presenting its mission and work. An example for a similar organisation that is part of the Commission is the European Centre for Algorithmic Transparency.

Hosting the office within the Commission allows making use of existing resources and expertise, notably through synergies with the structures built up to enforce other digital files, such as the European Centre for Algorithmic Transparency, and synergies with related initiatives at EU level, such as the EuroHPC Joint Undertaking and the AI Testing and Experimentation Facilities under the Digital Europe Programme for tasks like auditing and testing.

In addition, hosting the Office within the Commission would allow the most effective use of resources, because it allows making use of existing administrative structures, avoiding the additional cost that would be linked to the set-up of a new entity. The additional staffing needs could be financed through levies from the very capable foundation models and GPAI used at scale.

Section II – Prohibitions, law enforcement and national security

To reconcile the positions of the co-legislators, this note proposes a package: to limit the additional prohibitions proposed by the Parliament and to consider those taken out of the list of prohibitions high-risk, whilst possibly adding safeguards such as a monitoring and an oversight mechanism at EU level.

A. PROHIBITIONS

1. Real-time biometric identification for law enforcement in publicly accessible spaces

The Parliament proposes to prohibit real-time remote biometric identification in publicly accessible spaces (RBI) for any purpose, without any exception**.** The Council on the other hand broadened the exceptions to the prohibition the Commission had initially proposed. A possible compromise could consist in keeping the prohibition for real time RBI but subject it to narrower exceptions and to add additional safeguards:

1) **Narrowing down the exceptions** compared to the initial Commission proposal (and Council GA):
   (i) **Victims of crime:** The Commission proposal foresaw the use of RBI for the search for victims of all crimes. This exception could be limited to the targeted search for victims of abduction, trafficking in human beings and sexual exploitation of women and children.
   (ii) **Prevention of crimes:** while the Council broadened this exception from the Commission proposal,[2] a compromise could be to revert to the initial Commission proposal: the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or of a terrorist attack. Moreover the protection of critical infrastructure could be limited to situations of serious disturbances, for example of energy; water, gas supply leading to a damage for an important number of citizens.
   (iii) **Prosecution of crimes:** The Commission proposed the possibility to use RBI for all 32 crimes referred to for the European Arrest Warrant. The exception could be restricted to only 16 most serious crimes[3] out of this list.

---

2 Changes introduced by the Council in its general approach: (ii) the prevention of a specific~~, substantial and imminent~~ ***and substantial*** threat to the ***critical infrastructure, life, health***~~life~~ or physical safety of natural persons or ~~of a~~***the prevention of*** terrorist ~~attack~~***attacks***;

3 Annex XXX (proposal of reduction of crimes from the JHA list):
- participation in a criminal organisation involved in one or more crimes listed hereinafter
- terrorism,
- trafficking in human beings,
- sexual exploitation of women and children and child pornography,
- illicit trafficking in narcotic drugs and psychotropic substances,
- illicit trafficking in weapons, munitions and explosives,
- murder, grievous bodily injury,
- illicit trade in human organs and tissue,
- kidnapping, illegal restraint and hostage-taking,
- racism and xenophobia,
- organised or armed robbery,
- illicit trafficking in nuclear or radioactive materials,
- rape,

2) **Adding safeguards:** Possible safeguards, cumulative or alternative, could include:
   a) To oblige Member States to notify to the Commission national legislation implementing the exceptions at the latest 30 days following the adoption of such legislation. Commission will check its compliance with EU law.
   b) Use of real-time RBI only allowed on the basis of a judicial decision (instead of decision by a judicial authority or by an independent administrative authority as proposed by the Commission).
   c) Notify the individual use of the system to the national market surveillance authority (for law enforcement this is the sectoral supervisory authority or the data protection authority).
   d) Transparency based on a regular reporting by Member States to the Office or another institution/agency at European level:
      i. The reporting would be based on a template provided by the Office or another institution/agency at European level and could occur, for example twice a year.
      ii. It would cover all judicial decisions, approvals and rejections and include details on operation of the system (for how long is the system operating and how often, for what reason, number of persons on the watch list, estimate how many persons were concerned and how successfully the systems was operating/error rate).
      iii. Member States could be obliged to send aggregated data.
      iv. The Office or another institution/agency at European level could also publish annual reports (with aggregated data) for the public with due regard for the protection of sensitive operational data.
   e) Oversight: In its role as guardian of the treaties, the Commission will monitor the implementation of the regime (in particular the correct use of the exemptions and the safeguards). The Commission could act upon complaints of citizens and civil society organisations or investigate ex-officio, supported by the Office and based on the regular reporting by Member States (as described above). If problems or abuses are identified, the Commission can engage with individual Member States. It can launch a formal infringement procedure, if it has sufficient reasons to consider systematic abuse or incorrect implementation of Article 5(1)(d)(-x) and no action has been taken by the Member State to correct it. The Commission role should be without prejudice to the DPAs roles to enforce the data protection rules. The DPAs have to act in complete independence pursuant to Article 8(3) of the Charter.
   f) Recall in a recital there are complaint and judicial remedies at national level under the Data Protection legislation and a complaint mechanism under the AI Act [4]. The national market surveillance authorities under the AI Act can call upon the assistance of the AI Office.

2. 'Post' remote biometric data identification for law enforcement purposes

---

- arson,
- crimes within the jurisdiction of the International Criminal Court,
- sabotage

[4] EP and Council have already agreed to include possibilities for complaints to national market surveillance authorities in article 68a. As the negotiations stand, the AI Act does not include an explicit right to a judicial remedy, although such a right was proposed by EP but Member States considered it unnecessary as already applicable under national law. Under the EU Charter and under national law people should have such a right. It is important to note that for remote biometric identification rights to complaint and judicial remedy under data protection law would also apply as far as processing of personal data is concerned.

The EP wants to allow post remote biometric identification for law enforcement purposes only after a judicial authorisation for the targeted search connected to a specific criminal offense as defined in Article 83 TFEU[5]. The Council agreed with the Commission approach to consider these systems 'high risk', and added some additional exceptions for law enforcement (and border management) (see below under section D.

A possible compromise may include:

a. No authorisation needed for initial generalised checks and controls (relating to a concrete crime but not to "following" an individual or for the first identification of crime perpetrators.
b. Authorisation by independent administrative or judicial authority: for targeted search and retrieval of information for a specific identified individual from multiple publicly accessible spaces (investigations targeted at an individual which probably already may require a judiciary authorisation under national law).

Another possible compromise could be that the authorisation described in point b would be substituted a notification of the individual use of the system to the national market surveillance authority (for law enforcement this is the sectoral supervisory authority or the data protection authority).

3. Emotion recognition

EP wants to prohibit emotion recognition in law enforcement, border management, workplace and education institutions. We could explore 2 solutions:

Solution 1: A limited prohibition for specific use cases in these four sectors as listed in Annex III of the AI Act could be explored. As part of a balanced compromise, such a prohibition would require as a minimum:

a. To clearly define emotion recognition as being targeted at individuals and excluding screening of groups/crowds.
b. To include certain exemptions from the prohibition, notably for authorised medical and safety reasons (including detecting safety-related fatigue and drowsiness) and for possibly beneficial use cases.

Solution 2: A limited prohibition for specific use cases only in workplace and education institutions could be envisaged, leaving emotion recognition in law enforcement and border management as high risk, as in the Council's mandate.

4. Biometric categorisation based on protected data

The Parliament proposes a prohibition of biometric categorisation based on protected data, except for approved therapeutical purposes. The processing of such data in principle is already prohibited

[5] Article 83 TFEU only refers to crimes with a cross-border dimension, including corruption but excluding murder.

under data protection rules, unless strictly necessary, subject to appropriate safeguards and, in the field of law enforcement, where authorised under Member States or Union law. Currently protected data include protected characteristics under EU non-discrimination law (e.g. race, ethnic origin, age, sex, sexual orientation, religious beliefs, political opinions etc. as well as special categories of personal data protected under EU data protection law (e.g. personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs; trade-union membership; genetic data, biometric data processed solely to identify a human being; health-related data; data concerning a person's sex life or sexual orientation).

Biometric categorisation is an important tool for law enforcement authorities. As a compromise, a possible prohibition could be limited only to political opinions, religious beliefs, and sexual orientation unless -in the context of law enforcement- those characteristics have a direct link with a specific crime or threat (e.g. politically motivated crime or religiously motivated hate crime). In this context, the AI Act would specify the rules on the processing of biometric data contained in the Law Enforcement Directive, subjecting such use to the additional requirements set out in the AIA.

5. Prohibition of untargeted scraping of facial images to build facial recognition databases

EP wants to prohibit AI systems that create or expand facial recognition databases through the untargeted scraping of facial images from the internet or CCTV footage. At the level of the WP Telecom delegations indicated openness to the possibility of adding this prohibition.

6. Individual predictive policing for criminal and administrative offences

The EP proposes to prohibit AI system for making risk assessments of natural persons or groups thereof in order to assess the risk of a natural person for offending or reoffending or for predicting the occurrence or reoccurrence of an actual or potential criminal or administrative offence. Again, predictive policing is considered to be an important tool for an effective work of the law enforcement authorities.

A proposed compromise solution could be to include a prohibition under the social scoring prohibition in a more targeted way. This should focus on cases where fundamental/civil rights and freedoms are restricted in the context of law enforcement activities based solely on the AI output, without human supervision, predicting a behaviour of a natural person that constitutes criminal offence[6].

B. CONTROVERSIAL HIGH-RISK USE CASES

1.Biometrics

[6] For example, one could add to article 5(1)c) an additional element *iii) restrictions of fundamental/civil rights and freedoms in the context of law enforcement activities based solely on the AI output predicting a behaviour of a natural person that constitutes criminal offence without a reasonable suspicion.*

a. The Parliament extends the use case to all biometric identification systems (except those for authentication/verification and except for getting access to building, service etc., but only if it is 'one to one' identification). The Parliament also does not limit the use case to remote biometric identification (which normally occurs at a distance and under uncertain conditions, for example the screening of a metro entrance). This means that all biometric identification systems where there is comparison with a database ('one to many') would fall under the high-risk category, even if the person is actively participating, for example when giving fingerprints. The compromise could be the re-integration of the concept of 'remote', thus allowing to exclude "one to many identification" for reasons of access to buildings form the high-risk category (similar to Council and Commission's original proposal).

b. The Parliament proposes to define as high risk all systems that infer data based on biometric and biometric-based data[7] (e.g. biometric categorisation/ emotion recognition in so far not prohibited).
Agreement could be explored to include emotion recognition and biometric categorisation as high-risk, insofar these categories are being removed from the list of prohibitions. As part of the safeguards, it could also be explored to subject these new use cases to third party conformity assessment (similar to what the Commission has proposed for remote biometric identification systems). It would be also important to restrict any compromise solution to biometric data only (without inclusion of biometric-based data).

2. Law enforcement

a. The EP proposes to prohibit (see above) individual risk assessments and profiling of individuals for predictive policing
This would interlink with the conditions described in point 6 in section A above. For the rest it should remain high-risk (as originally proposed by the Commission).

b. Crime analytics (big data) related to individuals (originally proposed by the Commission)
Council proposed to delete the use case.

A deletion could be part of a balanced compromise, since the impact of crime analytics on fundamental rights is lower in comparison to predictive policing and profiling.

3. Migration and border control

a. Verification of authenticity of travel documents (originally proposed by Commission)
Council proposed to delete the use case. A deletion could be part of a balanced compromise, since systems only confirm the authenticity of travel documents (forgery) and in view of the increasing quality of these systems do not pose a particular high risk to fundamental rights.

b. The Parliament adds a use case for detecting and identifying natural persons in border management activities
This use case could be better targeted to protect rights of individual migrants, excluding the verification of travel documents and establishing a person's identity.

c. The Parliament adds a use case for forecasting of trends in migration and border-crossing.

[7] EP definition of biometric-based data is aligned with the Council broader definition of biometric data as both approaches go beyond GDPR that limits biometric data only for identifying individuals.

As part of a balanced compromise, this use case could be deleted.

For points b. and c. it appears to be a balanced compromise to accept point b, where there could be an impact on fundamental rights, and to reject point c., where the impact on fundamental rights appears to be negligible.

C. LAW ENFORCEMENT EXCEPTIONS

A balanced compromise could entail accepting all exceptions for law enforcement authorities introduced by the Council[8] except for the following:

1. **Article 51(1):** Council excludes law enforcement and border control providers and users from the registration in the EU public database for high-risk AI systems.
   Rather than a complete exception, providers and users from law enforcement/border management could have to provide only limited information in the public database[9]. Alternatively, providers and users from law enforcement/border management could be required to perform registration but in a non-publicly available part of the database.
2. **Article 83(1**): The Commission proposed not to apply in principle the regulation to AI systems components of large-scale IT systems in the Justice and Home Affairs area that have been put into service prior to 12 months after the date of application of the AI Act (i.e. initially 2 years after entry into force, still under negotiation). These AI systems would need to comply with the AI Act only if the legal acts establishing the large-scale IT systems are amended in a way that would lead to a significant change in the design or intended purpose of the AI systems components. The Parliament proposes to remove that exception and to require all AI systems part of large-scale IT systems to comply with the AI Act by 4 years after its entry into force.
   A compromise could be to keep the exception as formulated in the Commission proposal, while putting a sunset deadline, by which AI systems components of the large-scale IT systems need to be in any case made compliant, that is sufficiently long not to jeopardise their operation.

D. NATIONAL SECURITY EXCEPTION

The Council introduces an exemption for national security.

A balanced compromise could entail simplified wording closer to the Treaty: This Regulation is without prejudice to the competences and responsibility of the Member States with regard to their activities concerning defence and national security, in line with EU law.

Another possibility would be to keep the exemption for defence as in the Council's General Approach, but introduce more flexible wording with regard to national security, e.g. based on the wording from the Data Act. Additionally, the relevant case law could be referred to in the recitals.

---

8 For example, Article 29 (4), Article 61 (2) and Article 70(2) that excludes sensitive operational data;
Article 47 (1a) derogation from conformity assessment procedure in case of urgency; Article 52(2) derogation from the obligation to inform people about emotion recognition and biometric categorisation; Article 54, legal basis for further data processing in the sandbox.

9 See for example Washington law on biometrics that also require public accountability reports for law enforcement authorities and other public agencies.